A Work Project, presented as part of the requirements for the Award of a Masters Degree in Finance from NOVA School of Business and Economics

PATTERNS IN FINANCIAL MARKETS:

DYNAMIC TIME WARPING

MARIANA SÁTIRO COELHO, 319

A Project carried out with the supervision of

Pedro Lameira

January 2011

ABSTRACT

This work project introduces the performance of the algorithm Dynamic Time Warping amidst trading strategies in the financial markets. The employed procedure allows the comparison between any two sequences of data with different time lengths. Different features for the method were implemented, although those did not improve its promptness or accuracy in the outcomes obtained. Two potential investment strategies are presented within this theme. One yielded satisfactory outcomes whilst the other resulted in inconsistent values. The results point to the possible existence of patterns in the Equity Indexes' behaviour, as well as their distortion across the time axis.

Keywords: Dynamic Time Warping, Stock Pattern, Trading strategy.

The financial markets evolve throughout time and, depending on numerous global factors, their reaction pace is variable.

According to Chang, Liu et al (2009), the prediction of a stock includes the comprehension of two distinct features: market factors and unknown random processes. So, when one is considering a trading strategy, or even in any study of the markets, there are a considerable amount of factors that can be considered. However, the factor *Time* should not be left on the side, especially when the some of the investors' decisions are built on historical performances.

For example, when an economic factor is released, based on its expectations, the investors will take a determined position, which will lead to movements in stocks prices. At a certain point of the financial markets' life, an announcement such as the one previously referred to can cause an impact within a short time frame, such as a week. Conversely, if one is in the presence of a high volume trading period, the same news may possibly have influence for only an hour.

So, the adjustment of time is undeniable. Now the question is how to quantify and possibly integrate it into the economic studies, in the case that this even proves to be possible.

The main goal of this essay is to substantiate the existence of certain patterns in the financial markets, which can, or not, suffer from a time distortion. Then, one is eager to perceive behaviours on financial stocks which can be detected after being adjusted through a time modulation. Therefore, in case of satisfactory results, the work developed may be used as a way of prediction or even designed to be an investment strategy, as it can be seen later in the present report. Furthermore, there might be space to final remarks regarding the time's complexity in this matter.

Consequently, so that one can establish some possible trends, the need to determine a proper measure of similarity between two periods has shown to be crucial. For that purpose, the algorithm **Dynamic Time Warping** (DTW) is used, which along with some pre chosen liquid indexes, will assist in the gathering of the numerical results.

The mathematical procedure can be translated into a succession of steps that results in the comparison of two different sequences. This was suggested by Vintsyuk (1968) embedded in the area of speech recognition, which is until today the foremost applied topic of this method. It appeared as a solution for the traditional measure, the Euclidean distance.

However, the idea behind it had already been developed since the 40's, when Richard Bellman, a remarkable mathematician, proposed a direct relationship between optimization and programming. He faced the need of taking into account the factor *Time* for procedures that learn by themselves across time, denoting the concept by Dynamic Programming. Additionally, one should underline that machine auto learning methods have become a common practise today and it is used among the traders (Chang, Liu et al, 2009).

There has been a considerable amount of authors developing the process of Dynamic Time Warping within the context of Speech Recognition, such as Munich and Perona (1999). The core support of the idea that has been subject to several analysis is that one individual can say the same word in 30 seconds or in 1 minute, suffering therefore from a time distortion. Though, the word is still the same and consequently the sequences can possibly admit a match between them, enabling the discovery of the general pattern for that specific word.

Two of the utmost relevant researchers were Sakoe and Chiba (1978), who predominantly worked in the area of Speech Recognition. Both authors contributed significantly to the method's improvement, obtaining successful answers and results. Some of their suggestions are applied in the present paper.

Following the same criteria that had been employed in spoken languages before, Brown, Hodgings-Davis and Miller (2006) addressed a new theme: killer whale sounds. In this paper, it is affirmed that with the proper adaptation of the method, it is conceivable to preview if a whale might or will have dangerous behaviours based on the sounds that it produces.

One should perceive the importance of this method in different types of science related to sounds, either for translation purposes, the security industry or general voice.

Nevertheless, the algorithm is not only harnessed to sounds, it can also be ascribed to movement. Pohl and Hadjakos (2010) showed how a certain movement, previously captured by sensors, was decoded to a time series format and then compared through DTW. This happened to produce interesting and useful outcomes. Such investigation may be useful in Computer Games or even to DJs' study. Boulgouris, Plataniotis, Hatzinakos (2004) did a similar exercise, although the purpose was in some way different: the development of a security system where a person is recognized by his walk, which once again is attached to time variations.

Besides, Dynamic Time Warping can be applied to non-variable time data, as drawings or sketches. Per example, Niels (2004), Bashir and Kempt (2008), Romero, Kragic, Kyrki and Argyros (2008) demonstrated how two written characters can be compared by DTW. The great use of such application can be observed in the verification of signatures or even to encode texts. From an analogous perspective, Kovacs-Vanja (2000) also used the same algorithm to relate fingerprints.

All the previously mentioned authors have contributed, in one way or another, with new techniques or slight nuances. Some of them have improved the prompt of the method or simply provided an adaption to specific sets of data, which supports the flexibility of the process.

On a scale of the Financial Markets, the algorithm has been tested by using distinctive perspectives. Banavas, Denham and Denham (2000) proposed a model, including also the minimum embedding segment dimension method, which allowed them to select some trends. Wong and Yeung (2008) provided a forecast to particular stocks and then its correlation was computed with the real stock price. The outcome was quite satisfactory. The same subject has also been tackled by Chang, Liu et al (2009). A highly developed computing structure that learned through time was implemented. The results were incredibly favourable, whereas there are some issues regarding the over fitting of the method.

The present paper tries to apply the different versions of pure Dynamic Time Warping to financial indexes, without any previous treatment of data or combination of further techniques to improve the algorithm's performance. The method's advantages and weaknesses will be highlighted, so that it might be possible to understand whether or not the algorithm can be useful in the Financial Markets, either for studying or for prediction. The consistent results accomplished in a considerable range of study areas suggest that there is a good chance for it being applied in different subjects.

The paper is structured so that, in the first part a further description of Dynamic Time Warping is provided. It then proceeds to a detailed explanation of the two possible uses of the method as an investment strategy. Finally, both the results and the conclusions are presented.

METHODOLOGY

1. The algorithm: Dynamic Time Warping

Dynamic Time Warping (DTW) is a mathematical method that allows the comparison of two arrays of data. As it has been previously mentioned, it is applied in several areas, mostly in time series. However, it is also possible to operate it on static data, such as fingerprints. For every two subsequences of data, the algorithm gives us, not only the information of how alike they are, but also the best correspondence among their data prints. Furthermore, the procedure is quite flexible, so it can be easily adjusted to the type of data to which one intends to operate in.

Moreover, the most important aspect of this procedure is that it comprises the fact that the sequences may endure different durations. Therefore, it tolerates compression and distension of time. So, sequences of one week and one month can be easily confronted.

Consider two sets of points, A and B, with lengths of respectively n and $m \in \mathbb{N}$. The DTW starts to measure the distance from all the points of A to all elements in B. When one proceeds throughout this computation, there are some distance functions that can be used, for example the Manhattan Distance[1], the ρ-norm Distance[2] or even the Discrete Metric[3]. As it was suggested by Wong and Yeung (2008), in the present report one will apply the Euclidean Distance, which is a particular case of the ρ-norm Distance. One must highlight that that the DTW is more accurate than the regular Euclidean distance, since it returns a nonlinear alignment between sequences with different lengths (Figure 1 and Figure 5).

The Euclidean Distance can be represented by the following formula:

$$d_{i,j} = \sqrt{(a_i - b_j)^2}\text{, where } i \in \{1,2,\dots,n\},\ j \in \{1,2,\dots,m\},\ a \in A, b \in B \qquad \textbf{(1)}$$

[1]The Manhattan Distance can be expressed as $MD(\alpha,\beta) = \sum_{i=1}^{n} |\alpha_i - \beta_i|, \forall \alpha,\beta \in \mathbb{R}^n$

[2] The ρ-norm Distance is defined by $\rho(\alpha,\beta) = \sqrt[p]{\sum_{i=1}^{n} |\alpha_i - \beta_i|^p}, \forall \alpha,\beta \in \mathbb{R}^n$

[3] The Discrete Metric is given by $MD(\alpha,\beta) = \begin{cases} 0, \alpha \neq \beta \\ 1, \alpha = \beta \end{cases}, \forall \alpha,\beta \in \mathbb{R}^n$

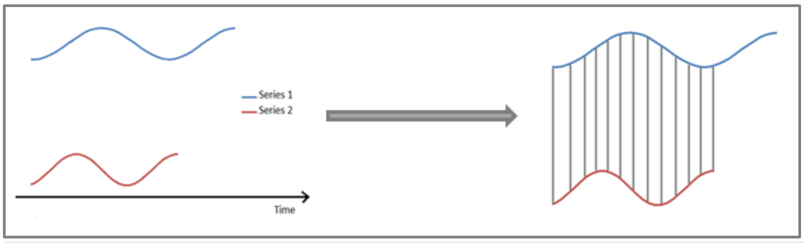

Figure 1 – Euclidean Measure.
For two distinct series, the Euclidean measure returns a straightforward comparison. Points with the same index are connected.

For instance, choose the second point of the set A, a_2, and then calculate the distance to all the other points in B (Figure 2). Then the same procedure is repeated to the rest of the points in the first sequence. This route results in a matrix that can enclose the following aspect: $\begin{bmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nm} \end{bmatrix}$. Remark that the entrance of the matrix (i, j)stands for the formula **(1)** applied to the point *i* in A and *j* in B.

Afterwards, the accumulated cost matrix is computed, following the rule to each cell:

$$c_{i,j} = \begin{cases} d_{i,j} & \text{if } i = 1 \wedge j = 1 \\ d_{i,j} + min\{d_{i,j-1}, d_{i-1,j}, d_{i-1,j-1}\} & otherwise \end{cases} \quad \textbf{(2)}$$

The element *(i,j)* of the previous matrix stands for the accumulated cost of inserting the correspondence between the points a_i with b_j in the final match. So, the value represents the cost of that correspondence, *(i,j)* , plus the minimum backward correspondences in the considered path. This matrix is essential for the next step: the search for the minimum path in the accumulated cost matrix, which coincides with the best match between the sequences.

For the calculation of the optimal match, one can assume that the first elements of the sequences are connected, as well as the last ones, in order to obtain a reasonable correspondence (Figure 3). Henceforth, the path must start at the entrance (1,1) and finish in (n, m). Finally, the path is simply computed by continuously searching the smallest value in the neighbourhood of the matrix. So, supposing that one is in the cell (i, j) in the procedure, it means that, in order to have the best alignment, the point *i* of A must be connected to the element j of B, the next point to become part of the minimum path coincides with the lowest value in the set $G = \{d_{i,j+1}, d_{i+1,j}, d_{i+1,j+1}\}$. The cells that are not a part of the final match are left with zeros.

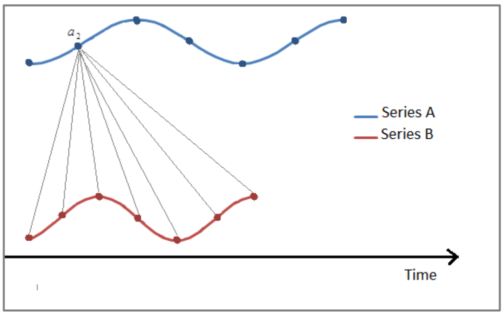

Figure 2 – Step of the DTW procedure:
The Euclidean distance is computed from a point of A to all points in B

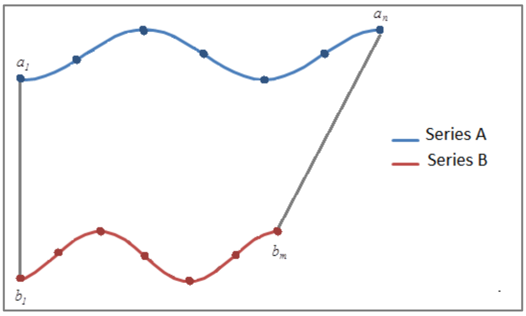

Figure 3 – Boundary Condition:
Independently of the length of the sequences, the firsts and lasts elements are connected

The total cost of the path corresponds to the sum of the cells of the match found, which is never lower than 0. The smaller the value returned by the distance, the more alike the two sequences ought to be.

One should note that some restrictions are required, so that the result encompasses a logical output. All of them have been already referred. The Boundary Condition reassures that the point $(1,1)$ and (n, m) are in the path. The Monotonicity Condition and the Step Size Condition were implied in the set G, by guaranteeing that each cell that is added to the path, results in a movement in the sequence and that each point has a correspondence on the other series (Müller, 2007).

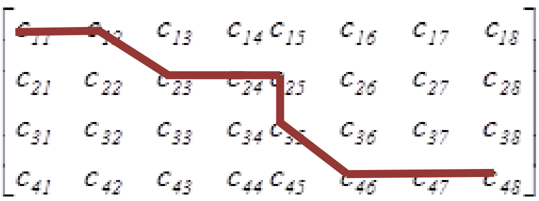

Figure 4 – Optimal Path in an accumulated cost matrix.
Possible optimal path between two sequences of lengths of 4 and 8

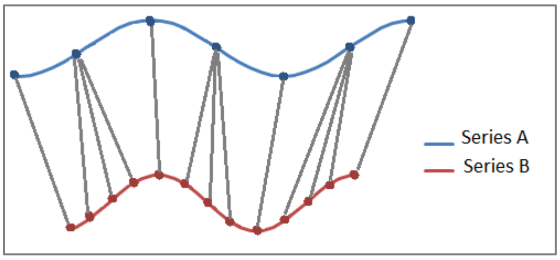

Figure 5- Optimal match between two series.
Each element has at least a connection.

2. Investment strategy

In this report two different investment strategies are assessed. They are here referred to as **Rolling** strategy and **Search and Invest** strategy. In both of them, the Dynamic Time Warping is employed as a way of learning with the purpose of investing in the subsequent period. Consequently, the original series is divided into two parts: the *on part*, in which the algorithm is going to learn the patterns and study its development, and the *off part*, in which the knowledge acquired is going to be tested. In both strategies there are certain shared nuances that can be modified according to the goal that the user specifies.

When the algorithm is applied, it is possible to adapt the type of matching that one is allowing for. For instance, if the user decides that the match cannot take extreme correspondences, an **extra boundary condition** is added. In the calculation of the minimum match, the corners of the matrix are not considered. This is reflected in the already mentioned set G, which is modified to $G_p = \{d_{i,j+1}, d_{i+1,j}, d_{i+1,j+1} : i+1 \le n-p \wedge j+1 \le m-p\}$, where p stands for the length that one is willing to ignore. (Figure 6)

Moreover, in the computation of the minimum path, the research neighbourhood may be extended. That is controlled through the **slope** of each match in the matrix. In that situation, the set G can be slightly adapted to $G_{id,jd} = \{d_{i+k,j+f} : (1 \le k \le id) \wedge (1 \le f \le jd)\}$. In an extreme case, the slope equals to $s = \frac{id}{jd}$, where *id* and *jd* stand respectively for the vertical and horizontal length allowed (Figure 7). In the set $G_{id,jd}$ the slope is variable; however it is also possible to fasten it. Both features were suggested by Sakoe and Chiba (1978).

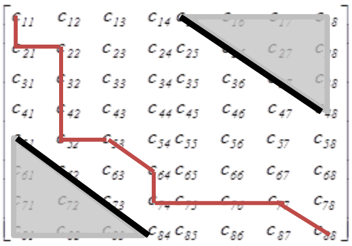

Figure 6 – Optimal path for p=4
The corners of the matrix are not considered. Extreme matches are not allowed.

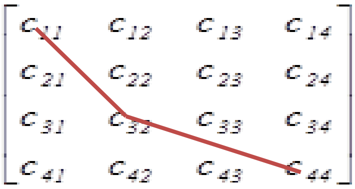

Figure 7 – Optimal path matrix for variable slope
The neighbourhood of searching is augment. In this case not every element of the series has a correspondence.

Regarding the investments' characteristics, the number of days that one wishes to **hold each position** (h) is also manipulated, as well as the **minimum expected return** (r) that the user wants to impose so as to invest. Finally, since the procedure's groundwork is the comparison among subsequences of data, certain characteristics of the comparison, as the **length of the subsamples** and the **number of steps** that concede the procedure to advance to the next comparison, are also variable. One must also highlight that the window movement is the most relevant point in this implementation.

Therefore, there is a set of key parameters that are easily controlled so as to better perceive the outputs and its sensitivity to changes. (In Appendices, Table 3)

a) <u>Rolling Strategy</u>

In this first strategy, the last *n* days of the *on part* of the data are chosen. Afterwards, it starts to compare those *n* days to the rest of the data in the *on part*. The best matches founded are considered. Then, a position is taken, short or long, based on what happened in the past data after the matched period, subject to the expected return being equal or higher than the minimum expected return ®. Finally, the *on part* is extended and the last *n* days are different, so the process is continuously repeated (Figure 8 and 9).

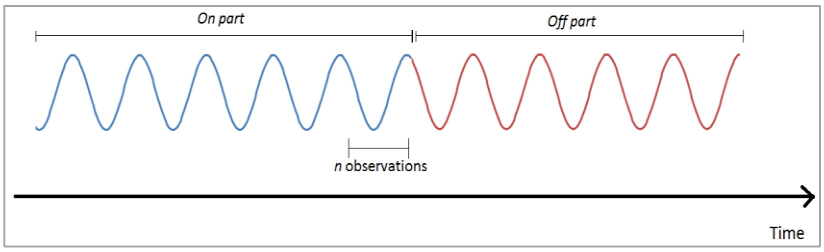

Figure 8 – Division of the series to execute the Rolling Strategy
The series is divided into two parts: *on* and *off*. From the *on part*, the last *n* observations are chosen .The *off part* is where the strategy is going to be tested.

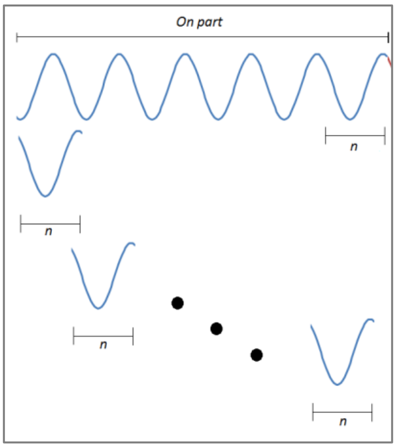

Figure 9 – The comparison process
On the right, it is presented a scheme of how the comparison is processed. The last sample of observation is taken and then it is compare against past periods.

Consider that one is comparing two sequences of different lengths, supposing *w* and *z*, being *w* higher than *z*. If the holding period was set to *s* days, a correspondence must be done in order to select the observed holding period to establish a decision. So the correct number of observations to study the behaviour of the series with length of w would be $s * \frac{w}{z}$.

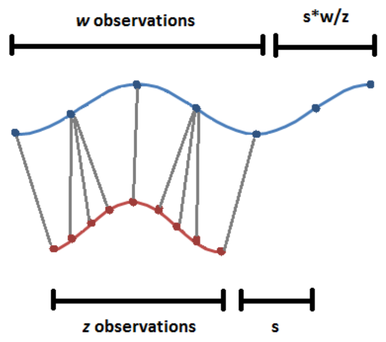

Figure 10 – Choosing the investment When the sequences have different lengths an adjustment ought to be made

Additionally, other features, such as volume and volatility can be also included. Observe the rationality of such hypothesis: the aim is to pursuit whether there are patterns in the financial markets that suffer from time distortion or not; However there are several other factors that can help with the characterization of a period, such as the volatility of the markets – which indicates the presence of turbulent or calm seasons, and the trading volume- since the number of investors and the amount traded is utterly relevant, due to the fact that it reflects the importance and impact of each price observation in the stock movement.

A similar approach was already used by Wong and Yeung (2008), however the DTW was applied distinctly to volume, volatility and returns and then a trend was inferred through the composition of the three distinct results. In the present report, it is suggested that the DTW should be applied only once, by transforming the formula **(1)** into:

$$d_{i,j} = \sqrt{(r_i - r_j)^2 + (v_i - v_j)^2 + (vu_i - vu_j)^2}\ , \quad \textbf{(3)}$$

where $i \in \{1,2,\dots,n\}$, $j \in \{1,2,\dots,m\}$, $a \in A, b \in B$ and v and vu stand for the volatility and volume respectively.

Ultimately, four different versions of the **Rolling** strategy were implemented: the basic, which only uses the returns, one that includes the volatility, another one which takes volume into account and the last one that incorporates all the three characteristics.

b) <u>Search and Invest Strategy</u>

The second approach foresees the investment from a different perspective. In the *on part* one chose X sequences, which are the potential patterns. Those were selected as being the sequences that were more frequent, throughout the *on part*. Their time length is a parameter that the user can specify.

Subsequently, in the *off part*, one is going to verify whether those patterns repeat themselves or not. When one compares each pattern to the observations in the *off part*, one is searching for similar behaviours, so that an investment position can be taken. Therefore, the comparison is only made for the first observations of the patterns, since the last part, which has a time length that equals the holding period, will define the investment decision.

So, by choosing one from the X sequences, one starts to compare it with the off part, using the DTW. When there is a sequence of the series that exhibits a certain degree of similarity (predefined) with the pattern, one will invest, according to the last observations of the pattern, long or short.

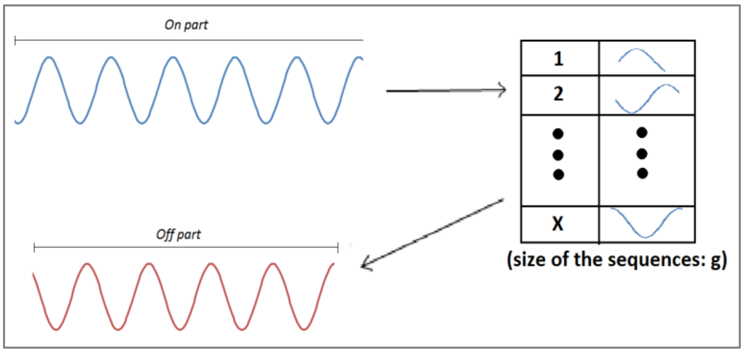

Figure 11 – Search and Invest Strategy
From the *on part* one selects X sequences which own the proper characteristics to become a good example of patterns (length=g). Then the first h observations of each sequence are selected, where h=g minus the holding days period. Later, it is tried to find behaviours in the *off part* that present a determined resemblance with any of the X sequences.

In summary, the parameters that both algorithms accept are: the size of each of the two sequences that are being confronted; the number of observations to advance for the next comparison; the size of the *on part*; the length of time that the user is keen to hold the position when one is taken; the minimum return that he requires; the number of patterns used for prediction and finally how the algorithm is going to find the minimum path, translated in the constraints *id*, *jd* and *p*, hitherto described. When volatility is considered, an extra feature is added: the amount of observations that are used to compute it.

Data

So as to test the hypothesis, 10 years of daily data of different stocks and stock indexes were used, all comprised in the period from January 2000 to December 2010, and reflected in approximately 3000 observations to each one. Nevertheless, one is are going to focus on the most traded and liquid ones: the major stock indexes. These include the French, Spanish, Hong Kong, American, European, Brazilian and Tokyo Indexes (respectively CAC, IBEX, HSI, SPX, SX5E, IBOV and NKY). In the present report the results for SX5E Index are shown. However, some comparisons are made with the other results, so as to acquire more robust and solid conclusions.

The data was taken from Bloomberg. It includes the last price and respective volume traded. Then, the returns and annualized volatility were computed, based on the general formulas:

$$r_i = \log \frac{p_i}{p_{i-1}} \qquad \textbf{(4)}$$

$$\sigma_a = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(r_i - \bar{r})} * \sqrt{260} \qquad \textbf{(5)}$$

Remark the possibility of changing the length of time used to compute the volatility.

Both the volatility and the volume were normalized with the formula $xn_i = \frac{x_i - \bar{x}}{\sigma}$, where the $\bar{x}$ and σ respectively stand for the mean and the standard deviation of each set and x_i is an observation, which in this case can be a reference to volatility or volume.

Should we not perform this standardization and the weights of those parameters would be much more relevant in the computation of the differences than the returns in formula **(3)**.

Furthermore, so as to measure the accuracy of the trading strategies, one used the Info Sharpe (IS) measure. The IS reflects the relationship between the return of the investment and its risk, measured through volatility: $IS = \frac{r}{\sigma}$. The Info Sharpe ratio is higher than 1 when the return exceeds the risk taken. Generally, one will consider that an investment strategy is worthy to engage when its IS is at least 1.

The programming language used was *Matlab*, due to its simplicity in dealing with large sets of data and its flexibility in manipulating matrix forms. The algorithm was implemented in its different versions and then applied to the indexes' returns and its corresponding volumes. The inputs values were changed, in order to test the sensitivity of the different parameters to variations.

RESULTS AND DISCUSSION

The empirical procedure relies on testing the algorithms heretofore implemented.

As it has been already emphasized, both strategies are based on a considerable amount of parameters, which can influence the strategy's success. However, so that the procedure's behaviour can be better understood, one is going to focus mainly on the sizes of the two sequences in comparison. The lengths of the potential patterns and of the subseries are the most relevant parameters in the analysis of time distortion.

Number of observations in the *on part*	1500
Number of observations to advance to the next comparison	5
Number of observations to add in the *on part* when extended	5
Number of sequences which we want to consider	5
Holding days period	3
Minimum required return	0.001
id	1
jd	1
p	0
Number of observations to compute volatility	10

Table 1 - Base inputs

The procedures were run with the same inputs to every index discussed. As a result, the same number of outputs was generated to each Index.

In regard to the strategy previously named **Rolling**, one starts to affix values for some of the parameters (Table 1). Meanwhile, the constraints that stand for the sequences' sizes were changed.

Concerning the SX5E Index, one may present the outputs grouped in Figure 12. In the first column of the table, the Info Sharpe values are considerable low. This case matches for length values of 5 days for the pattern. The results tend to improve by increasing the parameter. Nevertheless, the best values are achieved when the two series confronted preserve lengths from 30 to 50, for the pattern, against 15 or 40 days, for the subseries. This specific behaviour was observed in all the Indexes, excluding the SPX and NKY, which the highest performances were detected for sequences' sizes of 40 and 45 respectively and high patterns' sizes.

In the case when the two series have the same size, no impressive results are reached. For patterns with a length of higher values, for example 50, the results are quite satisfactory, even when the series sequence length is lower. This occurrence may be justified by the fact that potential noise becomes less relevant and a clearer pattern is inferred.

Pattern / Sequence	5	10	15	20	25	30	35	40	45	50
5	1.21	0.17	-0.05	-0.37	-0.02	1.74	0.76	1.19	0.44	0.85
10	-0.83	-1.25	0.21	0.53	1.03	1.16	0.62	0.69	1.30	1.07
15	-0.81	0.25	1.11	-0.35	0.40	1.62	1.69	1.11	1.41	1.14
20	0.13	-1.51	-0.93	-0.03	-1.38	-1.36	-0.62	0.68	0.92	0.78
25	-1.28	1.52	-0.05	0.18	0.42	0.09	0.76	0.78	1.81	1.65
30	-1.01	-0.18	0.63	0.00	0.03	-0.68	0.29	-0.90	-0.53	-0.08
35	0.39	0.74	-0.06	0.19	-0.11	-0.32	-0.57	1.27	1.46	0.78
40	-0.79	0.07	-0.05	0.51	1.13	2.34	2.52	1.63	1.89	3.12
45	0.83	1.40	0.88	0.71	1.53	1.16	2.38	0.23	0.77	1.62
50	1.05	-0.16	0.00	0.59	0.62	1.44	-1.21	0.23	-0.70	-0.32

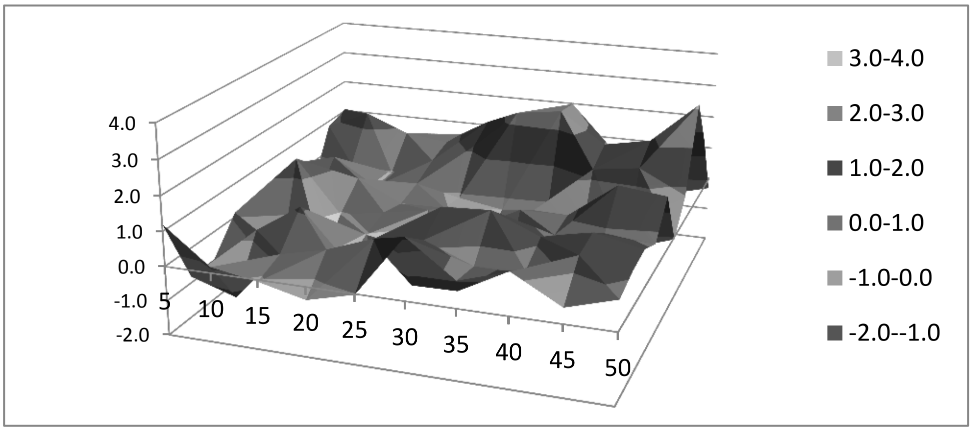

Figure 12 – Regular Rolling Strategy applied to SX5E Index
Using the base inputs and by varying the features related to size, the Info Sharpe's values were obtained. Reading the table throughout the horizontal and vertical axis, one analyse respectively the variation of pattern's and the sequence's size As for the graphic, the vertical axis stands for the Info Sharpe values that result from the execution of the algorithm for a pattern size represented in the scaled horizontal axis and for a sequences size in the last axis. The results were obtained for a minimum required return of 0.001.

The same outputs were also taken by changing the minimum required returns to 0.005 and 0.01 (in Appendices, Table 4). In the case of the SX5E Index, when the return was settled to 0.001, 31% of the Info Sharpe ratios from the table in Figure 12 were higher than 1; when one raises the required return to 0.005, the IS changes to 26% and for a minimum return of 0.01, the IS increases to 33%. With the exception for the SX5E and for the CAC Index, usually the Info Sharpe's values obtained with the regular **Rolling** strategy improved when the minimum required return decreased. However, if the algorithm yields low Info Sharpe for a minimum required return of 0.001, when the return is increased to 0.01 and 0.005, the output decreases even more. The mentioned percentage's values were roughly similar for all the Indexes, excepting for the SPX and NKY, which presented inferior results.

Table 2 summarizes the different versions of the Rolling strategy applied to the SX5E Index, as the comparisons among them. Including the volume or volatility in the calculations did not improve the results' accuracy. The regular Rolling strategy surpassed the other versions. From an overall perspective, this was verified in all Indexes, as it can be seen in the Appendices, where it is shown an average of the results obtained for the Rolling strategy applied to the mentioned Indexes (Table 5).

On average, the Rolling strategy allowed to invest approximately 28 times per year, which performs a rough value of 2 positions per month.

%	**Regular**			**Volume**			**Volatility**			**Volatility & Volume**		
	0.001	0.005	0.01	0.001	0.005	0.01	0.001	0.005	0.01	0.001	0.005	0.01
Positive Sharpes	67	59	59	52	57	75	50	49	44	53	51	52
Higher than 1	31	26	33	12	18	44	15	17	19	21	19	31
Higher than 2	4	8	14	0	1	21	4	6	8	2	1	10
Higher than Regular				37	43	55	39	42	45	37	42	48
Higher than volume	63	57	45				48	40	40	48	45	39
Higher than volatility	61	58	55	52	60	60				56	57	62
Higher than volatility & volume	63	58	52	52	55	61	44	43	38			

Table 2 – Summary of the outcomes for the Rolling Strategy applied to the SX5E Index
For each version of the **Rolling** strategy the chart presents the respective percentage of Info Sharpe values according to the descriptions. The inputs were the ones given in Table 1 with the exception for the minimum required return, which changes as observed. The last four rows provide a comparison between the different versions, also in terms of percentage.

A benchmark strategy was implemented in order to test the Rolling strategy against it. The considered benchmark, the trend following strategy, looks to the last observations, with a length equals to the holding days period of the Rolling Strategy, and if a trend is detected, up or down, an investment decision is taken, long or short respectively. The benchmark was tested to the 7 Indexes, taking into account the amount of times that the Rolling strategy allowed for an investment. When one compares the outcomes with the ones provided by the Rolling strategy, it is relevant to notice that the benchmark only exhibits a higher performance for the IBEX and the SPX Indexes.

Concerning the **Search and Invest** Strategy and according to Figure 13, one can observe that the Info Sharpe ratios are considerable high but highly inconsistent. This possibly occurred due to the fact that the amount of times that the algorithm allows for an investment is extremely low.

Pattern Sequence	5	10	15	20	25	30	35	40	45	50
5	-5.22	-0.86	4.55	3.36	-1.53	-5.20	-0.75	1.53	3.78	0.58
10	3.68	-0.98	0.54	-2.95	2.46	-1.49	-0.42	1.75	1.20	1.42
15	-0.36	2.79	-2.39	-1.39	0.91	1.05	2.21	3.15	2.05	-2.90
20	3.65	-3.87	5.73	-1.36	-1.06	0.63	1.16	3.30	0.25	-2.47
25	8.37	0.74	2.64	-0.67	3.81	-0.90	3.02	1.56	3.38	2.09
30	0.51	4.42	3.44	4.66	2.04	1.07	-1.72	-4.81	-4.00	-0.94
35	4.60	0.78	4.04	6.55	4.63	1.68	0.69	-2.27	-0.12	3.38
40	2.17	-0.59	0.19	-3.44	-1.63	2.73	-2.38	3.21	-4.03	1.64
45	2.49	0.33	0.00	0.28	0.30	3.54	3.54	1.20	-2.03	1.04
50	2.13	-3.81	-0.20	-1.16	4.10	-5.16	-0.14	2.76	6.09	1.37

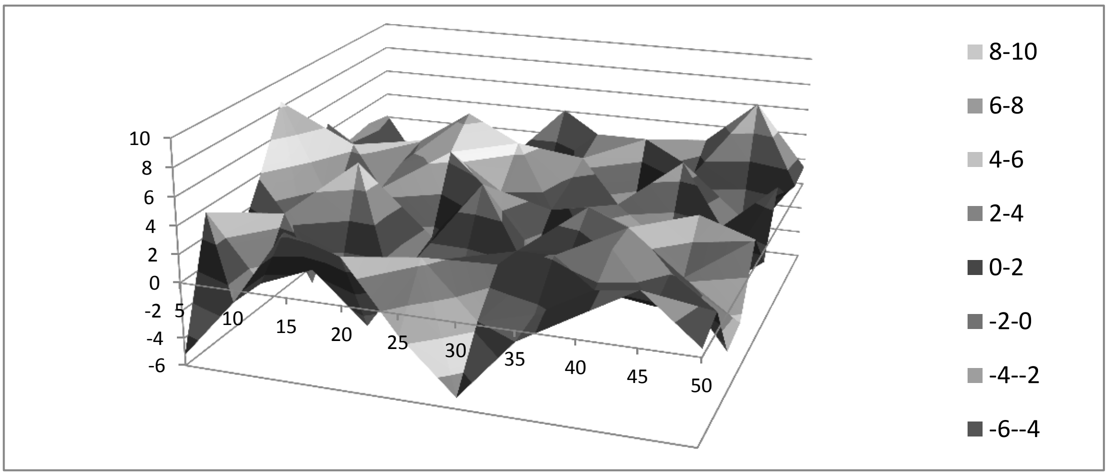

Figure 13 – Search and Invest Strategy applied to SX5E Index, for r = 0.001
By using the base inputs and by varying the features related to size, one obtained the Info Sharpe's values for the Search and Invest strategy. By reading the table throughout the horizontal and vertical axis, we analyse respectively the variation of the pattern's and the sequence's size

Once again, when the minimum required return is increased, in general it leads to an outcomes' improvement (Table 6). However, the SPX, the HSI and the SX5E Indexes presented some extreme cases under the latter strategy. For certain inputs, the method was not even able to establish any pattern and, henceforth, no investment was executed. The higher the minimum required return was, the more likely this was to occur. It is not possible to gather further conclusions that are consistent enough among most of the stocks.

As referred, the **Search and Invest** strategy only admits a profitable match nearly 6 times per year. This aspect is truly relevant, and together with the inconsistency of the results, it suggests that this strategy is more risky. Even though, it presents prominent Info Sharpe values. Therefore, there may be space to study it while applying it to a portfolio of indexes.

In order to evaluate the robustness of the results, a sensitivity analysis was performed. Having the values in Table 1 as a base model, one examined how the Info Sharpe ratios changed to variations in each parameter.

By varying the value of *p* (Sakoe and Chiba, 1978) from 0 to 40, which reflects the inclusion or not of the matrix's corners when searching for the optimal path, it is worth noticing that no significant changes were observed. This simply shows that in most of the cases, no extreme matches are made. This happened in all indexes tested with no exception and for all the different versions of the **Rolling** strategy.

The other feature suggested by Sakoe and Chiba (1978), the slope[4], also did not add substantial improvement to the results. By only varying one of the slope's characteristics, *id* or *jd*, the Info Sharpe ratios remained unchanged. For instance, for the CAC Index, when one limits the slope to values that satisfy $s = \frac{1}{jd}$ such that $5 \leq jd \leq 20$, the Info Sharpe is always equal to -0.815. Once again, this observation is valid for all the Indexes and for all the distinct

[4] One recalls the formula applied in the computation of the slope: $s = \frac{id}{jd}$

versions of the strategy. It is relevant to highlight that, while all the points in the sequences have a correspondence in the optimal match when calculated with the already mentioned set G, in the computation of the procedure with different slopes that possibly does not occur. However, the fact that the outcomes do not considerably change when we adjust the slope for valid values may indicate that the number of correspondences among the sequences' elements in comparison is not crucial in the search of the optimal match.

Similarly, for all the Indexes, the number of days used in the volatility's computation never presented any deviations when it is varied within the range 5 to 90 days. However, when combined with volume, the outcomes slightly increased with the range.

As it was expected, the features which are directly related to the comparison of the sequences, their lengths and the number of observations used to advance to the next comparisons, were the most relevant. In the case where one of those characteristics is altered, considerable changes are implied in the outputs. Although, it was also noticed that, in general, when greater results are achieved, translated to an Info Sharpe near or higher than 2, the results were found to be more robust to changes. Notwithstanding, the **Search and Invest** strategy is clearly the most volatile.

In summary, the version of the **Rolling** strategy that leads to more consistent outcomes is the one that considers volatility. However, the version that yielded higher results was the regular **Rolling** strategy. The **Search and Invest** strategy allowed for a scarcer number of investments, which is directly related to the amount of patterns found. A further study is then required to state firmer conclusions.

The Dynamic Time Warping algorithm appears to be valuable in the study of the financial markets. Not only it is a method that evolves through time, but it also allows for time's compression and distortion. Although, its time consumption is sizeable, its flexibility enables studies from distinct perspectives.

In accordance with the **Rolling** strategy, the outcomes suggested that it is possible to perceive patterns in the equity indexes analysed. Furthermore, those patterns seem to suffer from time distortions, as it is observable through the comparison between a pattern with 50 observations and a subseries of lower lengths.

Even though there are countless factors influencing the markets, *Time* should be also contemplated. Although investigations regarding this issue are extremely complex, and it is not totally possible to determine an exact measure, this characteristic can and should be scrutinized.

CONCLUSION AND FURTHER RESEARCHES

The present essay provides a base introductory study in the detection of patterns in the Financial Markets using the algorithm Dynamic Time Warping.

The algorithm DTW was enforced in various financial time series, in a data range from January 2000 to December 2010, where patterns were explored in order to, not only test the procedure as a way of investment, but also to scrutinize its efficiency. The Dynamic Time Warping method has the leading advantage of enclosing the adjustment of the time factor. Two different investment strategies were experimented. Both included particular details which could improve the quality of the match provided by the procedure. Those were related to algorithm itself - such as the slope, or to the investment - for example the holding days period. Its accuracy was measured by applying the Info Sharpe Ratio. Largely, the **Rolling** strategy accomplished satisfactory outcomes, namely when compared to a benchmark. The subsequent strategy tested, **Search and Invest**, reached some higher Info Sharpe values but a lot more inconsistent. In both of them, the features added, the slope and the type of match admitted, did not improve the outcomes. Therefore, those features ought to be relinquished.

Due to the massive dynamic portrayal in the current financial markets, this algorithm can prove to be valuable, however it must not be the unique tool employed since the results were not fully consistent and not as robust as one would like. For example, the SPX Index provided less profitable results within the implemented strategies. In addition, it is relevant to mention that if one excessively manipulates the inputs, it is possible to obtain over fitted outcomes with this method and therefore better results that may not correspond to the real values. However, with the appropriate precautions, it might be possible to take some profit out of a strategy that encompasses the Dynamic Time Warping algorithm.

As for futures researches, it can be advantageous to employ similar strategies building a portfolio of stocks, exploiting the diversification effect and therefore improving the portfolio's return. Moreover, diverse ways of investment can be tested and in particular the second investment strategy may be tested with more frequent observations. Considering patterns of any time length may prove to be advantageous, whereas it is computationally demanding. Finally, for a deeper study in this matter, the Dynamic Time Warping procedure can be combined with Hidden Markov Models (HMM), a stochastic process. The DTW and the HMM produce a non-linear sequence alignment, however there is a probability distribution attached to the HMM, which the DTW, a deterministic process, does not have. This method may reflect the other feature that Chang, Liu et al (2009) considered relevant in the market's study, the unknown random processes. The aggregation of both methods was already attempted in separate subjects (Fang, 2009). It is enclosed also in Dynamic Programming and can potentially capture other distinct features which DTW does not cover.

REFERENCES

[1] Banavas, G., Denham, S. and Denham, M., 2000, “Fast nonlinear deterministic forecasting of segmented stock indices using pattern matching and embedding techniques”, In *Computing in Economics and Finance 2000* 64, Society for Computational Economics.

[2] Bashir, M. and Kempt, J. 2008 “Reduced Dynamic Time Warping for Handwriting Recognition Based on Multi-dimensional Time Series of a Novel Pen Device”, In *World Academy of Science, Engineering and Technology* 45, Paris.

[3] Bellman, R and Kalaba, R. 1959 “On adaptive control processes”, In *Automatic Control, IRE Transaction* 4(2): 1-9.

[4] Berndt, D. and Clifford, J. 1994 “Using Dynamic Time Warping to Find Patterns in Time Series”, In *Workshop on Knowledge Discovery in Databases* (KDD-94), Seattle, Washington.

[5] Boulgouris, N., Plataniotis, K. and Hatzinakos, D. 2004 “Gait Recognition Using Dynamic Time Warping*”*, In *IEEE International Workshop on Multimedia Signal Processing*: 263-266, Siena, Italy.

[6] Brina, C., Niels, R. and Overvelde, A. DATA “Dynamic time warping: A new method in the study of poor handwriting” In *Journal of Human Movement Science* 27: 242-255.

[7] Brown, J., Hodgins-Davis, A. and Miller, P. 2006 “Classification of vocalizations of killer whales using dynamic time warping”, In *Journal of the Acoustical Society of America* 119: EL34-EL40.

[8] Carolyn, M., Gada, K., Martin, L., Keith, P., and Chris, S. 2000 “An Investigation of Machine Learning Based Prediction Systems”, In *The Journal of Systems and Software* 53:23–29.

[9] Chang, P., Liu, C., Lin, J., Fan, C. and Ng, C. 2009 “A neural network with a case based dynamic window for stock trading prediction”, In *Expert Systems with Application* 36: 6889-6898.

[10] Efrat, A., Fan, Q. and Venkatasubramanian, S. 2007 “Curve Matching, Time Warping, and Light Fields: New Algorithms for Computing Similarity between Curves”, In *Journal of Mathematical Imaging and Vision* 27: 203-216.

[11] Fang, C. 2009 “From Dynamic Time Warping (DTW) to Hidden Markov Model (HMM)”, *Final project report for ECE742 Stochastic Decision*, University of Cincinnati.

[12] Giorgino, T. 2009 “Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package” In *Journal of Statistical Software* 31(7): 1-24.

[13] Gudmundsson, S., Runarsson, T. and Sigurdsson, S. 2008 “Support Vector Machines and Dynamic Time Warping for Time Series”, In *IEEE International Joint Conference on Neural Networks*: 2772-2776.

[14] Junkui, L. and Yuanzhen, W. 2007 “Early abandon to accelerate exact dynamic time warping”, In *The International Arab Journal of Information Technology* 6(2): 144- 152.

[15] Júnior, S. 2007 “Dynamic Time Warping baseado na Transformada Wavelet”, *Dissertation for title of Master in Applied Physics*, Sao Paulo University.

[16] Keogh, E. 2002 “Exact indexing of Dynamic Time Warping”, In *28th International Conference on Very Large Data Bases*, Hong Kong.

[17] Keogh, E. Pazzani, M. 2001 “Derivative Dynamic Time Warping”, In *First SIAM International Conference on Data Mining*, Chicago, USA.

[18] Keogh, E. and Pazzani, M. 1999 “Scaling up Dynamic Time Warping to Massive Datasets” In *Proceeding of the Third European Conference on Principles and Practise of Knowledge Discovery in Databases*.

[19] Keogh, E. and Ratanamahatana, C. 2004 “Everything about Dynamic Time Warping is Wrong”, In *Third Workshop on Mining Temporal and Sequential Data, in conjunction with the Tenth ACM SICKDD International Conference on Knowledge Discovery and Data Mining*, Seatle, WA.

[20] Kovacs-Vanja, Z. 2000 “Fingerprint verification system based on Triangular Matching and Dynamic Time Warping”, In *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11): 1266-1276.

[21] Müller, M. 2007 *Information Retrieval for Music and Motion*. Springer: Chap 4.

[22] Munich, M. and Perona, P. 1999 “Continuous Dynamic Time Warping for translation-invariant curve alignment with applications to signature verification”, In *7th International Conference on Computer Vision*, Korfu, Greece.

[23] Nandyala, S. and Kumar, T. 2010 “Real Time Isolated Word Speech Recognition System for Human Computer Interaction”, In *International Journal of Computer Applications* 12(2): 1-7.

[24] Niels, R. 2004 "Dynamic Time Warping: An intuitive way of handwriting recognition?" Master's thesis, Radbound Unversity Nijmegen.

[25] Pohl, H. and Hadjakos, A. 2010 "Dance Pattern Recognition using Dynamic Time Warping" In *Proceeding of the 7th Sound and Music Computing Conference*.

[26] Romero, J., Kragic, D., Kyki, V. and Argyros, A. 2008 "Dynamic Time Warping for Binocular Hand Tracking and Reconstruction" In *2008 IEEE International Conference on Robotics and Autornation*.

[27] Sakoe, H. and Chiba, S. 1978 "Dynamic Time Warping Programming Algorithm Optimization for Spoken Word Recognition" In *IEEE Transactions on Acoustics, Speech and Signal Processing* 26(1): 43-49.

[28] Salvador, S. and Chan, P. 2007 "FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space" In *IEEE Transactions on Biomedical Engineering* 43(4).

[29] Shorten, G. and Burke, M. 2011 "The Application of Dynamic Time Warping to Measure the Accuracy of ECF Compression" In *The International Journal of Circuits, Systems and Signal Processing* 5: 305-313.

[30] Vintsyuk, T. 1968 "Speech Discrimination by Dynamic Programming" In *Kybernetika* 4: 81-88. English Translation.

[31] Wong, P. and Yeung, C. 2008 "A Novel Combined Approach for Time Series Data Mining: Stock Trend prediction Based on Matched Subsequences from Dynamic Time Warping in Wavelet Decomposed Time Series" *CMP510 Project*, Hong Kong, China.

APPENDICES

Parameter	Description	Related to
p	limitation in the number of the matrix cells when computing the optimal path	Calculation of the optimal path
id, jd	Slope of the match	
lp	length of the pattern	Comparison
ls	length of the subseries	
n	number of steps to advance in comparison	
h	Holding days period	Investment
r	Minimum expected return	

Table 3 – Key parameters of the investment strategies

Sequence \ Pattern	5	10	15	20	25	30	35	40	45	50
5	1.85	-0.02	-0.13	-1.25	0.40	0.86	0.04	-0.17	-0.24	-0.34
10	-0.92	-1.11	0.62	-0.23	0.99	1.35	0.15	-0.53	-0.70	-0.43
15	0.08	-0.31	0.77	-0.50	1.70	2.17	2.48	1.91	1.53	0.69
20	-0.21	-1.98	-0.80	0.46	-1.73	-1.13	-0.95	0.87	1.13	1.68
25	-0.50	2.68	-0.23	0.02	0.31	0.18	1.83	0.48	1.27	0.79
30	-1.19	-1.32	1.06	-0.35	0.08	-1.32	1.35	0.34	0.39	-0.21
35	0.58	0.67	-0.93	-0.25	-1.79	-1.56	-1.66	-0.02	0.64	0.54
40	-0.18	-0.34	0.27	0.24	2.92	2.26	1.77	2.23	2.56	4.10
45	0.62	0.74	0.82	1.35	1.72	0.31	1.10	-0.76	1.63	1.83
50	0.95	0.46	0.04	-0.17	0.23	1.74	-0.90	-0.70	-0.04	-0.21

Sequence \ Pattern	5	10	15	20	25	30	35	40	45	50
5	1.53	-1.61	-2.30	0.22	-3.48	-2.11	-1.73	0.62	-1.51	-0.14
10	-1.91	-1.87	0.10	-0.23	1.18	1.33	1.82	1.34	0.04	-1.56
15	0.07	0.09	1.84	-0.65	2.30	2.28	3.70	2.12	4.44	1.62
20	-3.30	-0.05	-0.82	0.79	-4.14	-4.55	-3.54	0.58	-2.52	0.77
25	0.40	2.56	-0.74	0.96	0.90	-0.84	-0.69	0.20	0.27	0.91
30	-1.11	-0.14	0.90	0.80	0.43	-1.65	1.32	-0.51	-2.02	-3.07
35	-0.05	1.73	-1.51	-0.79	-1.77	0.22	-0.14	1.44	1.55	1.91
40	1.47	-1.11	1.72	0.53	4.83	3.17	3.60	3.45	7.67	5.99
45	-0.98	0.68	1.71	1.64	1.30	2.14	-0.02	0.28	1.32	2.46
50	0.28	0.75	-0.08	-3.67	0.25	0.17	-1.32	-0.49	1.26	-0.95

Table 4 – Outcomes for the regular Rolling Strategy applied to the SX5E Index
The table on the left stands for the values in Table 1 employed for a minimum required return of 0.005 while the other coincides for a return of 0.01

%	Regular			Volume			Volatility			Volatility & Volume		
	0.001	0.005	0.01	0.001	0.005	0.01	0.001	0.005	0.01	0.001	0.005	0.01
Positive Sharpes	60	58	57	52	51	53	55	54	53	52	49	50
Higher than 1	25	30	32	15	23	36	18	22	19	18	22	27
Higher than 2	3	8	16	2	5	19	2	5	5	2	5	9
Higher than Regular				40	43	47	45	44	45	44	44	47
Higher than volume	60	57	53				52	49	49	53	51	50
Higher than volatility	55	56	55	48	51	51				46	46	50
Higher than volatility & vo	56	56	53	47	49	50	45	47	46			

Table 5 – Average Outcomes for the regular Rolling Strategy
The chart gives the average percentages of the **Rolling** versions. The results were obtained for all the Indexes and then a standard mean was computed. The first row stands for the percentage of Info Sharpe values higher than 1, while the following row present a comparison among all the different versions.

Sequence \ Pattern	5	10	15	20	25	30	35	40	45	50
5	-3.10	5.51	-0.79	-3.71	1.78	5.16	4.71	3.50	1.23	3.82
10	-7.40	0.47	-2.89	2.83	1.34	3.42	9.67	3.69	-0.07	3.04
15	-0.51	-6.32	2.44	-6.28	4.60	2.87	3.14	4.85	5.45	-1.27
20	-1.33	1.68	1.01	-0.17	-0.92	5.65	-0.34	4.34	-0.44	-0.80
25	-3.35	-0.73	2.21	-7.58	-0.01	1.05	3.38	7.06	9.93	3.17
30	-15.97	-0.73	-8.39	-16.75	-3.55	-0.18	2.56	3.12	4.91	5.40
35	-1.51	-0.93	-12.50	-11.39	7.11	-4.10	-0.05	1.35	6.21	2.32
40	1.84	-2.97	-5.94	-11.88	6.98	-7.78	-0.50	6.37	0.43	6.56
45	5.48	2.03	-4.70	3.31	15.32	5.55	8.63	-1.84	1.19	7.46
50	10.11	1.30	-0.94	6.37	2.44	2.19	0.86	-3.38	-7.79	1.56

Sequence \ Pattern	5	10	15	20	25	30	35	40	45	50
5	-2.34	6.81	-5.19	-3.04	2.22	5.41	3.18	6.24	3.50	9.55
10	-10.34	13.89	0	-2.80	-2.80	0.92	12.46	3.69	6.17	13.20
15	1.11	-8.51	0	0.46	-0.15	2.60	8.54	5.18	5.62	0.25
20	-5.55	1.22	0	-2.61	-3.94	0	0.56	0.82	1.79	-1.22
25	-4.22	-0.64	10.16	-3.88	-6.21	-1.71	21.01	8.68	9.97	7.99
30	-24.37	3.04	-2.15	-15.45	-3.55	-5.74	6.92	2.58	9.35	9.00
35	-1.65	-7.31	-9.39	-6.29	4.24	-5.90	-3.00	7.28	10.78	5.15
40	1.90	-2.46	-5.94	-6.29	6.87	-6.51	-7.13	2.76	2.67	5.34
45	12.48	0.61	-4.70	11.76	9.03	3.17	6.36	0.00	0.04	4.52
50	18.23	-8.02	-0.94	11.76	-1.64	5.89	-0.11	3.61	-7.14	1.65

Table 6 - Outcomes for the Search and Invest strategy applied to SX5EIndex
On the left is given the results for a minimum required return of 0.00**5** while the other table the ***r*** is changed to 0.01

Pattern Length	Rolling				Search and Invest
	Regullar	Volatilty	Volume	Volatility & Volume	
10	0.495	0.666	0.202	0.105	1.775
15	-0.965	0.191	-0.156	0.297	0.828
20	0.242	-0.300	-1.295	0.001	-1.591
25	0.448	-0.146	2.144	2.501	-0.051
30	-0.813	1.328	2.010	1.125	-1.011
35	-0.425	0.035	-0.775	-0.600	-6.399
40	-1.248	-0.832	-0.064	0.227	-0.570
45	-0.106	-1.169	-0.570	-1.000	-1.199
50	-0.318	-1.516	1.213	1.811	4.067

Series Length	Rolling				Search and Invest
	Regullar	Volatilty	Volume	Volatility & Volume	
10	0.574	1.219	1.993	1.066	-3.121
15	0.300	1.420	0.675	1.474	-0.527
20	0.304	0.657	1.428	0.635	-4.297
25	0.538	0.735	0.769	0.851	-0.106
30	-0.196	0.793	0.821	0.677	0.963
35	0.865	1.234	1.056	1.025	1.977
40	-0.301	1.303	1.055	1.789	-0.729
45	-0.829	0.588	1.276	1.337	-0.577
50	-0.104	0.917	1.332	1.422	0.240

Table 7 -– Sample of the outcomes of the sensitivity analyses for the size characteristics: pattern and subseries
In the charts can be found some of the Info Sharpe values obtained when changing only the length of the pattern (on the left) and the size of the subseries (on the right) in comparison. Both strategies were exposed to this analysis.